Proc. Natl. Acad. Sci. USA
Vol. 91, pp. 3092–3096, April 1994
Microbiology

G → A hypermutation of the human immunodeficiency virus type 1 genome: Evidence for dCTP pool imbalance during reverse transcription

(dNTP pools)

JEAN-PIERRE VARTANIAN*, ANDREAS MEYERHANS†, MONICA SALA*, AND SIMON WAIN-HOBSON*‡

*Unité de Rétrovirologie Moléculaire, Institut Pasteur, 75724 Paris, France; and †Abteilung Virologie, Institut für Medizinische Mikrobiologie und Hygiene, Universität Freiburg, 79104 Freiburg, Germany

Communicated by Manfred Eigen, December 28, 1993

ABSTRACT **The quasispecies model for RNA viruses predicts the existence of a replication error threshold beyond which there is a melting or total loss of sequence information. Retroviral G → A hypermutation is probably an example. Here it is shown that G → A transitions may occur in both GpG and GpA dinucleotide contexts. Transitions in GpG preferentially occur via base mispairing at the ends of runs of G residues, whereas G → A transitions within GpA may result from temporary dislocation of the primer and template strands by a single base. The two circumstances may be related by the local dCTP substrate concentration. An *in vitro* elongation assay shows that primer/template dislocation is more frequent for the human immunodeficiency virus type 1 reverse transcriptase than for murine or avian retroviral enzymes. Taken together these data suggest that G → A hypermutation is an example of induced mutation whereby the viral reverse transcriptase is forced into making errors by imbalances in the intracellular dCTP concentration.**

The concept of the viral quasispecies (1) was formulated with a view to describing the extensive genetic variation manifested by RNA viruses. Thus, a viral quasispecies may be understood as a discrete ensemble of viral genomes defined in time and space (2, 3). The human immunodeficiency viruses (HIVs) are no exception (4, 5). As the HIV provirus integrates into the host cell genome and may remain functionally silent, hence sequestered from immune responses, HIV variants are accumulated. Thus, there is extraordinary genetic diversity present, probably greater than for any other virus studied to date.

One of the features of the quasispecies model is the notion of a replication error threshold. A melting or total loss of sequence information is predicted if replication proceeds beyond this point (1). As the model has more than shown its utility it seems reasonable to ask what might such a melting of information correspond to in the real RNA viral world? Hypermutation of viral genomes immediately springs to mind (6–13). In the case of retroviruses, the G → A hypermutation is believed to occur during proviral synthesis (6, 7, 9–11) as opposed to the post-replication phenomenon of A → I hypermutation in measles virus (12, 13). Retroviral G → A hypermutation is erratic (S.W.-H., unpublished data), is particularly striking for the lentiviruses in contrast to other retroviruses (6, 7, 10, 11), and shows a distinct preference for the GpA dinucleotide (6). While not formally proven, it is believed to occur via repeated −1 dislocations of the nascent strand (6). However, the preference for the GpA dinucleotide might reflect a kinetic phenomenon and not strand slippage. As it cannot be reproduced by any experimental protocol to date the factors predisposing G → A hypermutation remain uncertain.

A balanced supply of the deoxyribonucleoside triphosphates (dNTPs) is essential to high-fidelity replication. The converse, decreased replication fidelity and asymmetric dNTP pools, is emphasized by a large body of experimental data (15–22). Recently it has been shown that perturbation of dNTP pools can either restrict or enhance HIV-1 replication. Restriction of HIV replication followed reduction of the intracellular dCTP pool (23). As this restriction was not completely overcome by addition of exogenous deoxycytidine, an augmented frequency of defective genomes could not be ruled out. Here data are provided that G → A hypermutation occurs in two dinucleotide contexts that may be related to fluctuations in the intracellular dCTP pool. It is shown by *in vitro* elongation studies that G → A substitutions within the GpA dinucleotide occur via strand slippage more frequently when HIV-1 reverse transcriptase (RT) is used as opposed to two other retroviral enzymes.

MATERIALS AND METHODS

Virus Samples. Those used were the HIV-1 B40 isolate passaged either on peripheral blood mononuclear cells (PBMCs) or on the U937-2 and Molt 3 cell lines (6, 24).

PCR, Cloning, and Sequencing. The 850-bp nef-U3/R region was amplified by PCR using NL1 and NL2 primers as described (25). PCR products were purified on a 1% low-melting point agarose gel, phosphorylated by T4 polynucleotide kinase, and ligated with *Sma* I-digested and dephosphorylated M13mp18 replicative form DNA. After transformation of *Escherichia coli* TG1, plaques were screened *in situ* by using a mixture of three oligonucleotides, S1, S2, and S3 (25). Twenty M13 recombinants from each of the PBMC, U937-2, and Molt 3 cultures were sequenced by the standard dideoxy method with dATP[α-^{35}S] (600 Ci/mmol; 1 Ci = 37 GBq) using M13 universal primer and the three oligonucleotides S1, S2, and S3. The samples were resolved on buffer gradient gels as described (6).

Subcloning, Transfection, and Assay for Soluble Alkaline Phosphatase (SEAP). The enhancer, promoter, and hypermutated TAR regions from clone 121 were amplified using the NL2 and NL3 primers (25). DNA was cleaved by *Hin*dIII and *Sal* I and ligated into pUC/SEAP (25). The validity of this construct, pTAR(G→A)/SEAP, was confirmed by double-stranded plasmid sequencing, using M13 reverse sequencing primer. Triplicate cultures of SW480 cells (human colon carcinoma cells) were cotransfected by pTAR(G→A)/SEAP

The publication costs of this article were defrayed in part by page charge payment. This article must therefore be hereby marked "*advertisement*" in accordance with 18 U.S.C. §1734 solely to indicate this fact.

Abbreviations: HIV, human immunodeficiency virus; PBMC, peripheral blood mononuclear cell; MoMLV, Moloney murine leukemia virus; AMV, avian myeloblastosis virus; RT, reverse transcriptase; SEAP, soluble alkaline phosphatase.

‡To whom reprint requests should be addressed.

and pSV2tat/LAI (4) by the calcium phosphate method. Sixty hours post-transfection culture supernatants were cleared by centrifugation (15,000 × *g*) and heated for 10 min at 65°C to inactivate any endogenous alkaline phosphatase. SEAP activities were determined using 50 μl of supernatant (26).

***In Vitro* Elongation Assay.** Experiments were performed with RTs from HIV-1 (27), Moloney murine leukemia virus (MoMLV; BRL), and avian myeloblastosis virus (AMV; Pharmacia). The DNA template was a 60-mer (5′-CTGAGGATCCATGTGGAATTCAGTGGGGAATAATCTAACTCTAGAGCACCAAGCTTCTGA) while the primer (18H, 5′-TCAGAAGCTTGGTGCTCT) was 18 bases long. The primer/template ratio was adjusted to 0.4 nM primer for 10 nM template (ratio 1:25) in labeling buffer (50 mM Tris·HCl, pH 7.6/10 mM $MgCl_2$/5 mM dithiothreitol/0.7 mM ATP). The hybrid was incubated with each of the three different RTs in optimized reaction buffers (28), in the absence of dCTP and an excess of dTTP—i.e., 0 μM dCTP, 250 μM dTTP, 25 μM dATP, and 25 μM dGTP. To favor template/primer dislocation, long elongation times of 1 min and 3 min at 37°C were used. This was followed by addition of 3 μl of prewarmed dNTP chase (250 μM dTTP, 25 μM dATP, 25 μM dGTP, and 1 mM dCTP) for 10 min at 37°C. The final concentration of primer/template was ≈0.8 ng/18 μl. As a positive control, elongation was performed using balanced dNTP concentrations (25 μM each). The reactions were stopped by incubation at 90°C for 5 min. The elongated products were PCR amplified for 15 cycles with the 18H and 18BE 5′-CTGAGGATCCATGTGGAA primers in standard conditions (6). Thermal cycling parameters were 95°C (15 sec), 55°C (15 sec), and 72°C (15 sec) with a final 10-min elongation step at 72°C. The products were purified on a 1.5% low-melting point agarose gel and cloned in frame into *Eco*RI and *Xba* I sites of M13mp18. Wild-type constructs could be distinguished from M13 phage by their light blue phenotype when plated with 0.1% 5-bromo-4-chloro-3-indolyl β-D-galactoside in the top agar. Recombinants harboring +1 and +2 frame shifts could be detected using the blue/white β-galactosidase assay. They were distinguished from the background by screening with the P+1 (5′-TTAGATTATTTCCCCACTGAA) and P+2 (5′-TTAGATTATTTTCCCCACTGAA) probes, specific for the +1 and +2 frameshift, respectively. Positive recombinants were confirmed by sequencing.

RESULTS AND DISCUSSION

Two Dinucleotide Contexts for G → A Transitions in Hypermutants. The nef/long terminal repeat region was amplified from the samples used in the original study on hypermutated HIV-1 genomes (6). Four clones out of 60 were found to be hypermutated (Fig. 1). Inspection of the sequence context showed a marked preference for both GpA and GpG, in contrast to our previous findings, where only a preference for GpA context was significant. For clone 77 the predominance of GpG → ApG transitions was particularly striking. It must be emphasized that this region, the promoter/TAR region, is more G-rich (31% G) than other parts of the HIV genome (24% G). However, the penchant for transitions within GpG was obvious even for the env–nef region, where the base composition was more typical of the HIV genome (27% G). The density of G → A transitions varied greatly between clones and appeared to be most erratic, clone 112 being a good example.

Hypermutation does not occur smoothly (Fig. 2). The same sequence could be mutated differently. Therefore the GpA and GpG contexts, while important, must reflect the consequences of a phenomenon and did not *per se* dictate substitutions. GpG → ApG transitions were most frequent at the end of a run of Gs. This suggests progressive depletion of the dCTP substrate and replacement by dTTP. Within the same clones G → A transitions within GpG could be found juxtaposed to those in GpA. The finding of a second dinucleotide preference does not overturn the original hypothesis that hypermutation followed from changes in the intracellular microenvironment, notably dCTP depletion (6). Rather it supports and extends the idea: partial or progressive depletion of dCTP would favor GpG → ApG transitions, especially at the end of runs of Gs, while dCTP exhaustion would favor GpA → ApA transitions. In the initial report G → A transitions were also found in GpG dinucleotides (GpA > GpG >

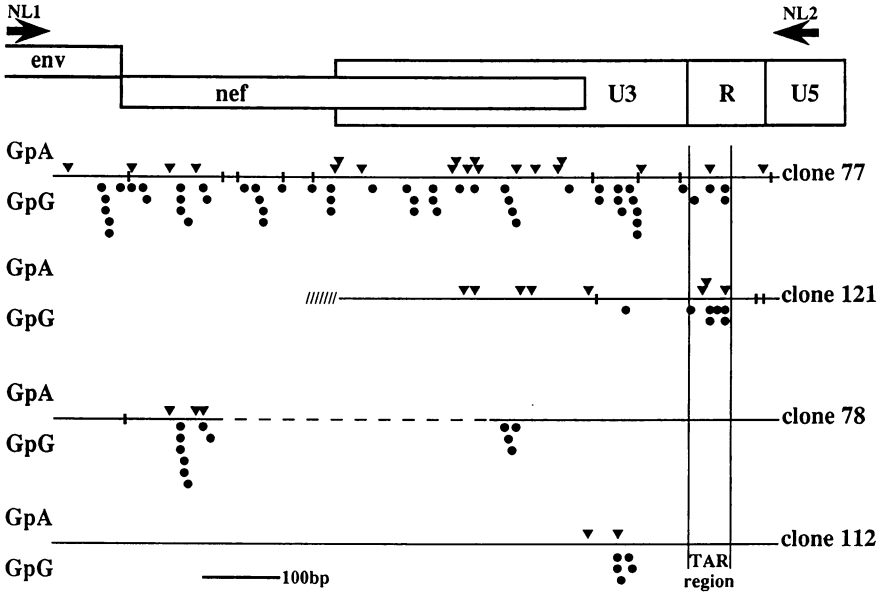

FIG. 1. Distribution of G → A transitions within four G → A hypermutants reveals two dinucleotide contexts. Solid triangles and circles represent G → A transitions within GpA and GpG dinucleotides, respectively. A vertical bar indicates a G → A transition within GpC or GpT. The TAR region, essential to efficient transcription, as well as the NL1 and NL2 amplification primers are shown. Clones 78 and 121 carried deletions. That in 121 occurred during the cloning process as the 5′ amplification oligonucleotide was missing. The deletion in clone 78 was typical of RT-generated deletions (7) and cannot be ascribed to PCR.

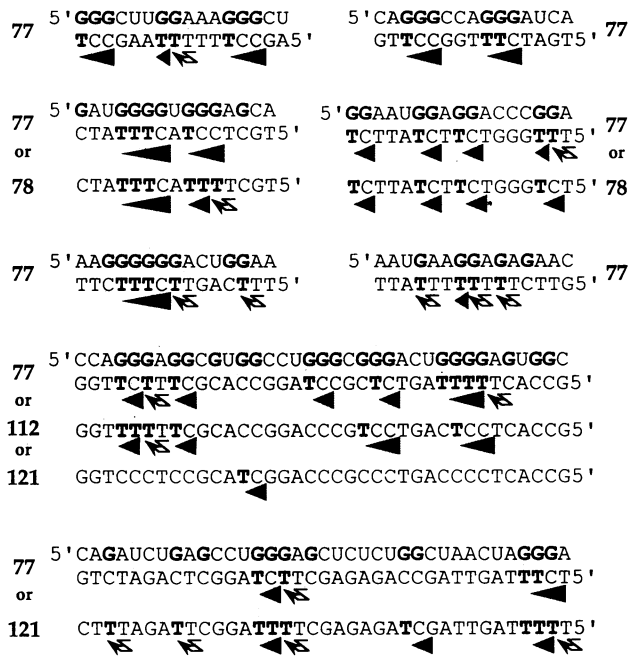

FIG. 2. G → A transitions within runs of G residues. A number of particularly striking examples are shown with respect to the plus strand reference sequence on top and hypermutated minus strand sequence underneath. Arrowheads emphasize the decreasing gradient of intracellular dCTP as polymerization progressed. The zigzagged arrow represents G → A transitions within GpA dinucleotides by dislocation mutagenesis. The striking differences between different hypermutants (clone 77, 78, 112, or 121) are evident. Boldface type indicates the G residues in the template and mismatched T residues.

GpT ~ GpC; ref. 6). As the frequency for GpG was similar to that expected from a random distribution, the finding could not be considered significant. The above data show that the proportion of G → A transitions in GpA and GpG can vary. That the two contexts were found interspersed could be taken to reflect the instability of the intracellular dCTP pool. Furthermore, the two dinucleotide contexts, as well as their idiosyncratic distribution, are probably the best evidence against invoking a mutant RT to explain the hypermutants. Like Maxwell's demon, it is difficult to conceive of an RT capable of identifying the last G in a run of Gs.

Hypermutated Promoter/TAR Regions Are Transcriptionally Silent. The finding of hypermutation on both strands is the clearest evidence that the phenomenon occurs during reverse transcription. However the marked asymmetry in the frequency of hypermutated viral plus and minus strands (6) might suggest a contribution from the cellular RNA polymerase II. To test this hypothesis the promoter/TAR region from clone 121 was cloned into the pUC/SEAP expression vector (25). This construct, pTAR(G→A)SEAP was transfected with the tat expressing plasmid, pSV2tat/LAI plasmid, into the SW480 cell line. As can be seen from Fig. 3*A*, the pTAR(G→A)SEAP construct was totally devoid of promoter activity. This finding excludes any contribution during the start of transcription. If hypermutation were to occur once transcription through the 5′ TAR region was completed, then only hypermutation in the U3 region and not the R region of the long terminal repeat should be observed (Fig. 3*B*). As the hypermutants harbored transitions in both the U3 and R regions, this result excludes any contribution to G → A hypermutation coming from aberrant transcription.

G → A Transitions May Occur via Dislocation Mutagenesis. G → A transitions within the GpA dinucleotide were suggested to occur via a mechanism termed dislocation mutagenesis (6, 29). However, a preference for G:T mismatchs

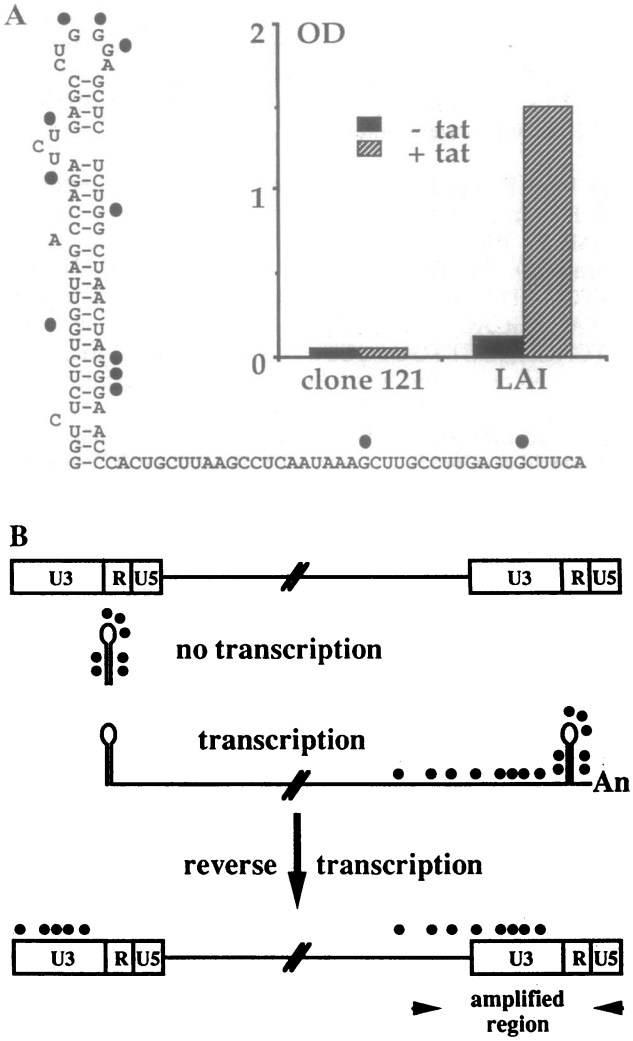

FIG. 3. Transcription does not contribute to G → A hypermutation. (*A*) The hypermutated U3/TAR region from clone 121 was transcriptionally silent when cloned upstream of an secreted alkaline phosphatase gene. (*B*) If hypermutation occurred after transcription was initiated—i.e., after TAR region synthesis—a hypermutated U3 region would be juxtaposed to a nonmutated R region. This is due to the rearrangement of the retroviral genome during provirus synthesis. That this is not observed rules out any contribution to G → A hypermutation from transcription.

within GpA might be preferred on kinetic grounds. To distinguish between the two, the following experimental strategy was developed. A 60-mer and a complementary 18-mer (18H) were synthesized and purified by HPLC (Fig. 4). Elongation from 18H by different RTs was carried out under highly asymmetric dNTP conditions (zero dCTP and excess dTTP), after which an excess of dCTP was added to capture dislocated intermediates. Because of the low primer/template concentrations used in the experiments, the products were then amplified by PCR. After digestion by *Eco*RI and *Xba* I, they were cloned in frame into the β-galactosidase reading frame of M13mp18 replicative form DNA. Recombinants could be identified by a light blue phenotype and were easily distinguished from wild-type M13. White plaques, which revealed frameshift mutants and cloning artifacts, were screened using specific oligonucleotide probes and sequenced.

As a control the reactions were carried out using balanced dNTP concentrations. This not only served as a reference point for each RT but also provided a control for the PCR amplification. For the AMV enzyme no frameshift mutants were identified (Table 1). For MoMLV the influence of

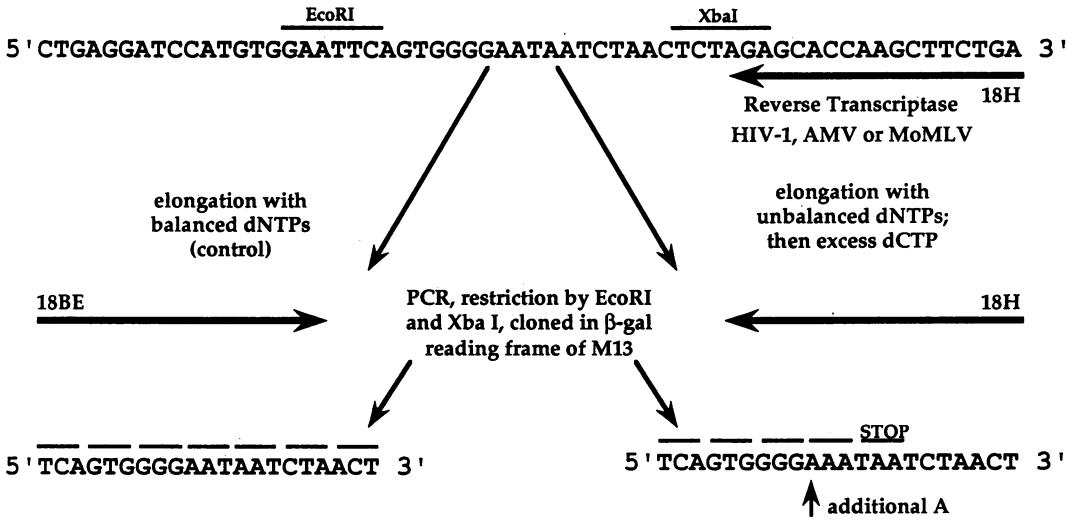

FIG. 4. Experimental protocol for trapping dislocated intermediates. The 60-mer incorporated internal *Eco*RI and *Xba* I restriction sites. From the GGGGA motif to the *Xba* I site the template was devoid of G residues allowing elongation by RT to start efficiently in the presence of dTTP, dATP, and dGTP alone. After 1 or 3 min an excess of dCTP was added and elongation was completed. The products were amplified by PCR (18BE and 18H) to allow molecular cloning via the *Eco*RI and *Xba* I sites. Recombinants harboring an extra A residue (↑), arising as a result of primer/template dislocation would change the reading frame of the β-galactosidase (β-gal) gene and terminate translation at the downstream stop codon. This results in a white phenotype when plated with isopropyl β-D-thiogalactoside and 5-bromo-4-chloro-3-indolyl β-D-galactoside. Other recombinants would be reflected by their light blue phenotype.

balanced or unbalanced dNTP concentration had no effect on the frequency of frameshift mutants. However, with HIV-1 RT, more frameshift mutants were identified when unbalanced dNTP pool was used (Table 1). Furthermore, the frequency increased 2-fold with longer elongation times, suggesting that dislocation of the primer with respect to the template was a slow phenomenon. In all cases only +1 frameshifts were identified.

Clearly, G → A transitions may occur via dislocation of the primer with respect to the template in the absence of dCTP. The degree of saturation of G by A may attain 60–80% for some segments of hypermutated lentiviral genomes (refs. 6 and 9–11; S.W.-H., unpublished data). For spleen necrosis virus, an avian oncovirus genetically more related to Mo-MLV than AMV or HIV-1, the proportion of substituted Gs was around 17% (7). The above finding, a higher frequency of captured dislocated intermediates when the HIV-1 RT was used, correlates well with a greater degree of saturation in lentiviral G → A hypermutated genomes.

It is clear from the above *in vitro* data and many other reports that the frequency of G → A transitions within GpA are of the order of 10^{-3} (15, 17, 28). How then is it possible to reconcile these observations with the finding of G → A hypermutation *in vivo*? The *in vitro* experiments are carried out with excess template/primer to determine the K_m and V_{max} of each reaction. *In vivo* the converse is true—i.e., an excess of enzyme over substrate [two genomes and probably of the order of 50–80 RT molecules per virion (14)]. Furthermore, RT is not a very processive enzyme and may dissociate from its primer/template substrate (28). Together these findings probably preclude the identification of G → A hypermutants *in vitro*, particularly when short elongation times, of the order of minutes, are used (15, 17, 28).

Taking together the data support the hypothesis that G → A hypermutation comes about as the result of a change in the local dCTP concentration. Low dCTP pools would favor transitions in GpG dinucleotides particularly at the end of runs of G. A dearth of dCTP would arrest elongation. Proviral synthesis may continue via dislocation of the primer with respect to the template in the context of GpA, GpT, and GpC. Dislocation within GpA would result in a relatively stable 5′ GpA/3′ TpT hybrid. Dislocation mutagenesis within GpT or GpC would lead to 5′ GpT/3′ ApA and 5′ GpC/3′ GpG mismatches. However elongation proceeds relatively unhindered only when dislocation occurs in the context of GpA as RT-mediated transitions are invariably more frequent than transversions (15, 17, 28). This helps explain why so few G → A transitions occur in the context of GpT and GpC.

This model abrogates the need for a mutant RT and reflects a large body of data associating dNTP pool imbalances, particularly dCTP reduction, with increased mutation. The highly erratic distribution of G → A transitions emphasizes the extraordinary sensitivity of reverse transcription to the intracellular dCTP concentration. In this sense the melting of sequence information predicted by the quasispecies model can be seen as being induced by the microenvironment within

Table 1. Frequency of primer/template dislocation in GpA during elongation by HIV-1, MoMLV, and AMV RTs

	HIV-1			MoMLV			AMV		
dNTP pool	*N*	*T*	Frequency	*N*	*T*	Frequency	*N*	*T*	Frequency
Balanced	3783	1	2.6×10^{-4}	8600	3	3.3×10^{-4}	3776	0	—
Unbalanced									
1 min	4801	3	6.2×10^{-4}	2864	1	3.4×10^{-4}	1363	0	—
3 min	2730	4	1.4×10^{-3}	900	0	—	1813	0	—

Unbalanced and balanced dNTP concentrations were 0 μM dCTP, 250 μM dTTP, 25 μM dATP, and 25 μM dGTP, and 25 μM dCTP, 25 μM dTTP, 25 μM dATP, and 25 μM dGTP, respectively. A number (*T*) of colonies having an additional base within a population (*N*) resulted from elongation when carried out using balanced as opposed to unbalanced dNTPs. Primer/template dislocation was more frequent when the HIV-1 RT was used as opposed to the MoMLV enzyme. Longer elongation times favored the formation of dislocated intermediates.

the porous nucleocapsid structure that makes up the viral replication complex.

This work was supported by grants from Institut Pasteur and l'Agence Nationale pour la Recherche sur le SIDA (J.-P.V. and S.W.-H.). A.M. was supported by grants from the Deutsche Forschungsgemeinschaft and Landesforschungsschwerpunkt des Landes Baden-Württemberg.

1. Eigen, M. & Biebricher, C. K. (1988) in *RNA Genetics*, eds. Domingo, E., Holland, J. J. & Ahlquist, P. (CRC, Boca Raton, FL), Vol. 3, pp. 211–245.
2. Steinhauer, D. A. & Holland, J. J. (1986) *Annu. Rev. Microbiol.* **41**, 409–433.
3. Domingo, E., Holland, J. J. & Ahlquist, P., eds. (1988) *RNA Genetics* (CRC, Boca Raton, FL), Vols. 1-3.
4. Meyerhans, A., Cheynier, R., Albert, J., Seth, M., Kwok, S., Sninsky, J., Morfeldt-Manson, L., Åsjö, B. & Wain-Hobson, S. (1989) *Cell* **58**, 901–910.
5. Wain-Hobson, S. (1992) *Curr. Top. Microbiol. Immunol.* **176**, 181–193.
6. Vartanian, J. P., Meyerhans, A., Åsjö, B. & Wain-Hobson, S. (1991) *J. Virol.* **65**, 1779–1788.
7. Pathak, V. K. & Temin, H. M. (1991) *Proc. Natl. Acad. Sci. USA* **87**, 6024–6028.
8. O'Hara, P. J., Nichol, S. T., Horodoyski, F. M. & Holland, J. J. (1984) *Cell* **36**, 915–924.
9. Johnson, P. R., Hamm, T. E., Goldstein, S., Kitov, S. & Hirsch, V. M. (1991) *Virology* **185**, 217–228.
10. Gao, F., Yue, F., White, A. T., Pappas, P. G., Barchue, J., Hanson, A. P., Greene, B. M., Sharp, P. M., Shaw, G. M. & Hahn, B. H. (1992) *Nature (London)* **359**, 495–499.
11. Perry, S., Flaherty, T., Kelly, M. T., Clabough, M. J., Tronick, D. L., Coggins, S. R., Whetter, L., Lengel, C. R. & Fuller, F. (1992) *J. Virol.* **66**, 4085–4097.
12. Cattaneo, R., Schmid, A., Eschle, D., Baczko, K., ter Meulen, V. & Billeter, M. A. (1988) *Cell* **55**, 255–265.
13. Bass, B. L., Weintraub, H., Cattenao, R. & Billeter, M. A. (1989) *Cell* **56**, 331.
14. Layne, S. P., Merges, M. J., Dembo, M., Spouge, J. L., Conley, S. R., Moore, J. P., Raina, J. L., Renz, H., Gelderblom, H. R. & Nara, P. L. (1992) *Virology* **189**, 695–714.
15. Bebenek, K., Roberts, J. D. & Kunkel, T. A. (1992) *J. Biol. Chem.* **267**, 3589–3596.
16. Kunkel, T. A. (1992) *BioEssays* **14**, 303–308.
17. Preston, B. D., Poiesz, B. J. & Loeb, L. A. (1988) *Science* **242**, 1168–1171.
18. Kunz, B. A. & Kohalmi, S. E. (1991) *Annu. Rev. Genet.* **25**, 339–359.
19. Mathews, C. K. & Ji, J. (1992) *BioEssays* **14**, 295–301.
20. Meuth, M. (1989) *Exp. Cell Res.* **181**, 305–316.
21. Phear, G., Nalbantoglu, J. & Meuth, M. (1987) *Proc. Natl. Acad. Sci. USA* **84**, 4450–4454.
22. Slabaugh, M. B., Howell, M. L., Wang, Y. & Mathews, C. K. (1991) *J. Virol.* **65**, 2290–2298.
23. Meyerhans, A., Vartanian, J. P., Hultgren, C., Plikat, U., Karlsson, A., Wang, L., Eriksson, S. & Wain-Hobson, S. (1994) *J. Virol.* **68**, 535–540.
24. Goodenow, M., Huet, T., Saurin, W., Kwok, S., Sninsky, J. & Wain-Hobson, S. (1989) *J. AIDS* **2**, 344–352.
25. Delassus, S., Cheynier, R. & Wain-Hobson, S. (1991) *J. Virol.* **65**, 225–231.
26. Berger, J., Hauber, J., Hauber, R., Geiger, R. & Cullen, B. R. (1988) *Gene* **66**, 1–10.
27. Hansen, S., Schultze, T. & Moelling, K. (1987) *J. Biol. Chem.* **262**, 12393–12396.
28. Ricchetti, M. & Buc, H. (1990) *EMBO J.* **9**, 1583–1593.
29. Kunkel, T. A. & Alexander, P. S. (1986) *J. Biol. Chem.* **261**, 160–166.